Correspondence

A family of DNA repair ligases in bacteria?

Geoffrey R. Weller, Aidan J. Doherty*

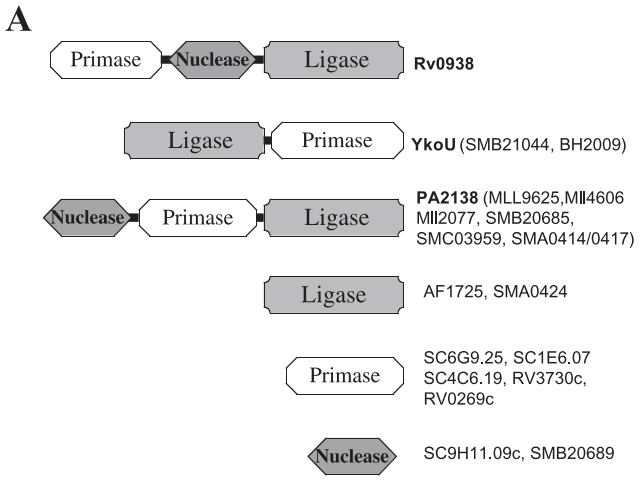
First published online 27 August 2001

The covalent rejoining of DNA ends at single-stranded or double-stranded DNA breaks is catalysed by DNA ligases. Ligases catalyse the formation of phosphodiester bonds at nick sites between 3′-hydroxyl and 5′-phosphate termini in double-stranded DNA [1]. DNA ligases are required for a number of important cellular processes including the replication of DNA, repair of damaged DNA, and for various recombination events within the cell. The ligase enzyme family can be divided into two broad classes: those requiring NAD^+ for activity and those requiring ATP [1]. The eukaryotic, viral and archaeal enzymes all require ATP, whereas the NAD^+-dependent ligases have only been found in bacteria. Despite their requirement for different co-factors both bacterial DNA ligases and their eukaryotic counterparts utilise a conserved catalytic mechanism [1]. Three DNA ligases (I, III and IV) have been identified in mammalian cells and it has been well documented that these distinct ligases play specific roles within the cell [2]. All eubacterial genomes encode an NAD^+-dependent ligase and, until recently, it was commonly believed that this enzyme was solely responsible for all DNA ligation reactions. However, it is now evident, from the sequences of many bacterial genomes, that some bacteria contain two or more DNA ligases, including ATP-dependent ligases. The presence of these additional ATP-dependent ligases suggests that prokaryotes, like eukaryotes, may employ specific ligases for defined roles in vivo, such as DNA repair and recombination.

We recently identified putative homologues of the Ku70 and Ku80 DNA repair proteins in many bacterial genomes [2]. Ku functions as a heterodimer in vivo that binds to the ends of DNA and promotes efficient and accurate DNA end-joining repair. Mammalian cells deficient in Ku are sensitive to ionising radiation. In our previous report we noted that many, but not all, of the bacterial Ku homologues are organised into operons containing putative ATP-dependent DNA ligases [2]. These include ligases encoded by YkoU (*Bacillus subtilis*), BH2209 (*Bacillus halodurans*), Rv0938 (*Mycobacterium tuberculosis*), PA2150 (*Pseudomonas aeruginosa*), and AFI1725 (*Archaeoglobus fulgidus*) (for a full sequence alignment of ATP-dependent ligases see www.ncbi.nlm.nih.gov/cgi-bin/COG/palox?COG1793). To detect related DNA ligases in other bacterial genomes we used YkoU and its orthologues as query sequences in BLAST searches of the protein sequence database. These searches retrieved additional putative ATP-dependent ligase orthologues (Mll2077, Mll4606, Mll9625 – *Mesorhizobium loti* and SMB20685, SMA0424 – *Sinorhizobium loti*) within operons containing Ku-like genes. We also noted the presence of previously undetected Ku-like genes in the genome of *S. loti* (SMb20686, SMb21406, SMb21407, SMa0426) and in many unfinished bacterial genomes including *Bacillus anthracis* (http://www.tigr.org/cgi-bin/BlastSearch/blast.cgi?organism=b_anthracis).

An examination of the secondary structure of the Ku-associated DNA ligases suggested that these putative proteins, with the exception of AF1725 and SMA0424, contain N- and C-terminal extension(s) attached to a catalytic core DNA ligase. We used these 'domains' as query sequences in BLAST searches of the protein sequence database in order to possibly predict additional functional domains. These searches were run to convergence with an *E* value of 0.01 and predicted, with high probability, the presence of two additional domains in many of the Ku-associated ligases. A domain, named lig primase, with significant homology to the small catalytic subunit of eukaryotic DNA primases, is present in many of the putative DNA ligases (Fig. 1A; for a full alignment see: www.ncbi.nlm.nih.gov/cgi-bin/COG/palox?COG3285). DNA primases are essential components of the DNA replication machinery in all organisms [3] and catalyse the synthesis of short stretches of RNA of defined length (9–12 bases) on single-strand DNA, acting as a primer for the replicative DNA polymerases. Eukaryotic and archaeal primases contain highly conserved sequence motifs (motif I and II) and sequence analysis showed that these motifs are present in the lig primase domains (Fig. 1B). Biochemical and structural analysis has shown that the two invariant Asp residues in motif I are essential for catalysis and, together with another conserved Asp residue, are likely to make up the catalytic triad [3,4]. Augustin et al. [4] recently reported that the conserved Asp residues of *Pfu* primase can be superimposed with a RMS deviation of 0.15 Å with the active site residues of four other polymerases. Eukaryotic, archaeal and lig primases share significant sequence homology with the family X polymerases [3], including pol β, an enzyme that only participates in DNA repair. However, primases have a different sequential arrangement of homologous residues and are likely to have evolved convergently to form their own distinct family of polymerases [4]. The lig primase domain, in common with eukaryotic primases, also contains the highly conserved motif II (Fig. 1B), implicated in the binding of both nucleoside triphosphates (NTP) and DNA [3,4]. Mutations of the conserved residues within this conserved motif in the mouse primase resulted in an increase in K_M indicating a role in nucleotide binding [3,4].

In addition to the primase domain, we detected the presence of another novel domain in a large subset of the putative ATP-dependent ligases (Fig. 1A,C). This domain, named lig nuclease, is typically fused to the N-terminal end of the ligase (Fig. 1A), with the exception of RV0938 where the domain is located between the primase and ligase domains (Fig. 1A). The BLAST analysis revealed that this domain has significant homology with the putative proteins SC9H11.09c and SMb20689, predicted to be related to the exonuclease III/AP (apurinic/apyrimidinic) DNA repair endonuclease family (see http://sequence.toulouse.inra.fr/meliloti.html). DNA damaging agents can produce base loss within DNA, forming AP sites or strand breaks with atypical 3′ termini. DNA repair at the AP sites is initiated by specific endonuclease cleavage of



B

	Motif I	Motif II
Sce	ELVF**D**ID**D**MDD	SG**RR**GAH
Spo	ELVF**D**ID**D**MID	SG**RR**GIH
Has	ELVF**D**ID**D**MID	SG**RR**GVH
Mmu	ELVF**D**ID**D**MID	SG**RR**GVH
Dme	ELVF**D**ID**D**MID	SG**RR**GIH
Afu	DLIF**D**ID**D**ADH	SGG**R**GYH
Mth	ELIF**D**V**D**AKD	SG-**R**GYH
Pho	ELVF**D**ID**D**AKD	SG-**R**GYH
Mja	ELAF**D**ID**D**VHK	SGN**R**GYH
BH	EIVF**D**LD**D**PPS	SGG**K**GLQ
BS	EIVI**D**LD**D**PPS	SGN**K**GIQ
ML	QIIF**D**LD**D**PDE	SGG**K**GYH
MT	RLVF**D**LD**D**PGE	SGS**K**GLH
PA	RFVL**D**LD**D**PDP	SGG**K**GMH

C

```
MLL2077    1-MPAPAKLKTYRAKREFSKTPEPAGGLISGD----GNRFVVHKHHATADHYDLRLQVGDVL
SMB20685   1--MATPKLEAYRKKRDFSKTPEPVGNLTAG-----GNRFVVHKHHATADHYDLRLEVDGVL
MLL9625    1--MASDKLTTYKAKRDFKKTQEPSGQGSVKPSNRRR--FVIQKHDATRLHYDLRLELGGVF
MLL4606    1----MAGLEQYHTKRDFKKTSEPAGKVARTGKNEAGGIFVIHKHAATRLHYDLRLEHDGVL
PA2138     1-MPSSKPLAEYARKRDFRQTPEPSGRKPRKDSTG-LLRYCVQKHDASRLHYDFRLELDGTL
RV0938   291-APVADRLTRYRRMRDASKTPEPIP--TAKPVTGDGNTFVIQEHHARRPHYDFRLECDGVL
                 *  *   *:  :* **             : :::* *   ***:**: ...:

MLL2077      KSWAVPRGPSLNPADKRLAVETEDHPLEYIDFEGVIPEGEYGGGPMIVWDTGTWAPMD--
SMB20685     KSWAVPKGPSLNPADKRLAVETEDHPLDYIDFEGVIPEGEYGGGPMIVWDKGVWAPMG--
MLL9625      KSWAVTRGPSLDPHDKRLAVEVEDHPLDYGDFEGTIPKGQYGGGTVMLWDRGYWEPEGKN
MLL4606      WSWAVTRGPSLDPHEKRLAVHVEDHPIDYAPFEGTIPKGEYGGGSVIVWDEGTWMPET--
PA2138       KSWAVPKGPCLDPAVKRLAVQVEDHPLDYADFEGSIPQGHYGAGDVIVWDRGAWTPLD--
RV0938       VSWAVPKNLPDNTSVNHLAIHTEDHPLEYATFEGAIPSGEYGAGKVIIWDSGTYDTEK--
              ****.:.   :.  ::**:..****::*  *** **.*.**.* :::** * : .

MLL2077      DVEKSLRTGAFKFRLAGEKLNGGWMLTRLKP-    145
SMB20685     DVDDDLRKGAFKFRLAGEKLNGGWMLAR-----   140
MLL9625      SPEQALAKGDFKFTLEGERLHGSF---------   141
MLL4606      DPAKAMKKGHISFELNGHKLHGLWHLVRLKPRAG- 149
PA2138       DPREGLEKGHLSFALDGEKLSGRWHLIRTNLRGR- 152
RV0938       -FHDDPHTGEVIVNLHGGRISGRYALIRTNGDRWL 440
                .   .* . . * * :: * :
```

Fig. 1. A: Domain organisation of putative ATP-dependent DNA ligases. The ligases are grouped according to domain organisation and the approximate location (amino acid) of individual domains is as follows: Rv0938 (primase, 1–280; nuclease, 281–460; ligase, 461–760), YkoU (ligase, 1–320; primase, 321–611) and PA2138 (nuclease, 1–200; primase, 201–550; ligase, 551–841). Protein accession numbers for other putative proteins containing related DNA ligase, primase and nuclease domains are also included. B: A sequence alignment of two catalytic motifs conserved across all of the eukaryotic-like DNA primases (small catalytic subunit) and the bacterial lig primase domains (bold). The organisms and the protein accession numbers are: Sce, *Saccharomyces cerevisiae* (P10363); Spo, *Schizosaccharomyces pombe* (O14215); Hsa, *Homo sapiens* (P49642); Mmu, *Mus musculus* (P20664); Dme, *Drosophila melanogaster* (Q24317); Afu, *Archaeoglobus fulgidus* (O29516); Mth, *Methanobacterium thermoautotrophicum* (O26685); Pho, *Pyrococcus horikoshii* (O57934); Mja, *Methanococcus jannaschii* (Q58249); BS, *Bacillus subtilis* (YkoU); BH, *Bacillus halodurans* (BH2209); ML, *Mesorhizobium loti* (Mll4606); MT, *Mycobacterium tuberculosis* (Rv0930) and PA, *Pseudomonas aeruginosa* (PA2150). C: A sequence alignment of a selection of the nuclease-like domains, lig nuclease, present in many ATP-dependent DNA ligases identified in PSI-BLAST searches. An asterisk indicates identical or conserved residues in all sequences in the alignment, a colon indicates conserved substitutions and a dot indicates semi-conserved substitutions. ←

the phosphodiester backbone. AP-like endonucleases are also generally capable of removing blocking groups from the 3′ terminus of DNA strand breaks. The lig nuclease domain also shares sequence homology with members of GIY-YIG (URI) nuclease family.

The crystal structures of ATP- and NAD^+-dependent ligases has revealed that these enzymes have a highly modular architecture consisting of a unique arrangement of two or more discrete domains [1]. Currently, five classes of domain (nucleotide binding domain, oligo binding fold, zinc finger, helix-hairpin-helix motif and BRCT domain) have been detected at both the sequence and structural level in the ligase super-family [1]. The additional domains of the larger ligases are thought to enhance DNA binding and targeting of the enzymes to sites of DNA damage, replication and recombination; however, none of these ancillary domains have known catalytic activities. In contrast, many of the putative ligases described in this report possess domains with significant similarity to proteins with known catalytic activities (primases and nucleases). The additional catalytic activities associated with these DNA ligases are, most likely, responsible for 'remodelling' damaged DNA prior to ligation. In many DNA repair systems, an endo/exonuclease activity is often required to resection damaged DNA, prior to later steps in the repair pathway. It will be of great interest to see if the lig nuclease domain possesses any nuclease-like activity. Recently, an archaeal primase was shown to preferentially synthesise DNA primers [5]. It is possible that the lig primase domain may be required for filling in short stretches of DNA at sites of DNA damage, preceding the final ligation step of repair.

The conserved organisation of Ku and ligase genes in the operons of many distinct bacterial genomes suggests a conservation of function in vivo. In mammalian cells, the Ku heterodimer complex interacts specifically with DNA ligase IV, recruiting the enzyme to sites of DNA damage [6]. Cells deficient in ligase IV are impaired in DNA double-strand break repair and are consequently hypersensitive to ionising radiation [7]. Similarly, we propose that in many eubacteria, Ku-like proteins may detect and bind to breaks in the genomic DNA and, subsequently, recruit ATP-dependent ligases to sites of DNA damage. Experiments are currently in progress to characterise the biochemical properties of these predicted ATP-dependent DNA ligases and also to establish the role played by these proteins in vivo.

Acknowledgements: A.J.D. is a Royal Society University Research Fellow and G.R.W. is supported by an MRC PhD studentship. Research in the A.J.D. laboratory is supported in part by the Royal Society, the Cancer Research Campaign and the Association for International Cancer research. We would like to thank Prof. S. Jackson and Dr R. Bowater for helpful discussions relating to this work. We apologise to authors whose work is not cited due to severe restrictions on references.

References

[1] Doherty, A.J. and Suh, S.W. (2000) Nucleic Acids Res. 28, 4051–4058.
[2] Doherty, A.J., Jackson, S.P. and Weller, G.R. (2001) FEBS Lett. 500, 186–188.
[3] Arezi, B. and Kuchta, R. (2000) Trends Biochem. Sci. 11, 572–576.
[4] Augustin, M.A., Huber, R. and Kaiser, J.T. (2001) Nature Struct. Biol. 1, 57–61.
[5] Bocquier, A.A., Liu, L., Cann, I.K., Komori, K., Kohda, D. and Ishino, Y. (2001) Curr. Biol. 11, 452–456.
[6] Nick McElhinny, S.A., Snowden, C.M., McCarville, J. and Ramsden, D.A. (2000) Mol. Cell. Biol. 20, 2996–3003.
[7] Riballo, E., Doherty, A.J., Dai, Y., Stiff, T., Oettinger, M.A., Jeggo, P.A. and Kysela, B. (2001) J. Biol. Chem. (in press).

*Corresponding author. Fax: (44)-1223-336827.
E-mail address: ajd42@cam.ac.uk (A.J. Doherty).

Cambridge Institute for Medical Research, Wellcome Trust Centre for the Molecular Mechanisms in Disease, and Department of Haematology, University of Cambridge, Hills Rd, Cambridge CB2 2XY, UK

PII: S0014-5793(01)02831-9